The Fractality of Polar and Reed-Muller Codes

Bernhard C. Geiger, *Member, IEEE*

***Abstract*—The generator matrices of polar codes and Reed-Muller codes are obtained by selecting rows from the Kronecker product of a lower-triangular binary square matrix. For polar codes, the selection is based on the Bhattacharyya parameter of the row, which is closely related to the error probability of the corresponding input bit under sequential decoding. For Reed-Muller codes, the selection is based on the Hamming weight of the row. This work investigates the properties of the index sets pointing to those rows in the infinite blocklength limit. In particular, the Lebesgue measure, the Hausdorff dimension, and the self-similarity of these sets will be discussed. It is shown that these index sets have several properties that are common to fractals.**

***Index Terms*—Polar codes, Reed-Muller codes, fractals, self-similarity**

I. Introduction

Polar codes and Reed-Muller codes are Kronecker product-based codes. Such a code of block-length 2^n is based on the n-fold Kronecker product $G(n) := F^{\otimes n}$, where

$$F := \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{1}$$

Following the terminology of [1], a rate-$K/2^n$ Kronecker product-based code is uniquely defined by a set $\mathcal{F}$ of K indices: Its generator matrix is the submatrix of $G(n)$ consisting of the rows indexed by $\mathcal{F}$. For polar codes [2], in which each row of $G(n)$ can be interpreted as a (partially polarized) channel, $\mathcal{F}$ consists of rows corresponding to the K channels with the lowest Bhattacharyya parameters [3] (the "good" channels, see Section II). For Reed-Muller codes, $\mathcal{F}$ consists of those rows of $G(n)$ with a Hamming weight above a certain threshold (see Section IV). Despite its importance for code construction, at least for polar codes, very little is known about the structure of $\mathcal{F}$. A recent exception is the work by Renes, Sutter, and Hassani, stating conditions under which polarized sets are aligned, i.e., under which the good (bad) channels derived from one binary-input memoryless channel are a subset of the good (bad) channels derived from another [4].

That Kronecker product-based codes, such as polar codes [2] or Reed-Muller codes, possess a fractal nature has been observed in [1], noting the similarity between $G(n)$ and the Sierpinski triangle. Much earlier, Abbe suspected that the set of "good" polarized channels is fractal [5]. Nevertheless, to the best of the author's knowledge, no definite statement regarding this fractal nature has been made yet. In this paper, we try to fill this gap and present results about the sets $\mathcal{F}$ for polar codes (Section III) and Reed-Muller codes (Section V). The self-similar structure of these sets is also suggested in [6], which shows that polar and Reed-Muller codes are decreasing monomial codes. While [6] focuses on finite blocklengths, we study the properties of $\mathcal{F}$ for infinite blocklengths, i.e., for $n \to \infty$.

To simplify analysis, we represent every infinite binary sequence indexed in $\mathcal{F}$ by a point in the unit interval $[0, 1]$. Let $\Omega := \{0, 1\}^\infty$ be the set of infinite binary sequences, and let $b := (b_1 b_2 \cdots) \in \Omega$ be an arbitrary such sequence. We abbreviate $b^n := (b_1 b_2 \cdots b_n)$. Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space with $\mathfrak{B}$ the Borel field generated by the cylinder sets $S(b^n) := \{w \in \Omega\colon w_1 = b_1, \dots, w_n = b_2\}$ and $\mathbb{P}$ a probability measure satisfying $\mathbb{P}(S(b^n)) = 1/2^n$. The following function $f\colon \Omega \to [0, 1]$ converts these sequences to real numbers:

$$f(b) := \sum_{n=1}^{\infty} \frac{b_n}{2^n} \tag{2}$$

Letting $\mathbb{D} := [0, 1] \cap \{p/2^n\colon p \in \mathbb{Z}, n \in \mathbb{N}\}$ denote the set of dyadic rationals in the unit interval, we recognize that f is non-injective:

Example 1. f maps both $b = (01111111 \cdots)$ and $b = (10000000 \cdots)$ to 0.5. We call the latter binary expansion *terminating*.

However, as the following lemma shows, f is bijective if we exclude the dyadic rationals:

Lemma 1 ([7, Exercises 7-10, p. 80])**.** *Let $\mathfrak{B}_{[0,1]}$ be the Borel σ-algebra on $[0, 1]$ and let λ be the Lebesgue measure. Then, the function f in* (2) *satisfies the following properties:*

1) *f is measurable w.r.t. $\mathfrak{B}_{[0,1]}$*
2) *f is bijective on $\Omega \setminus f^{-1}(\mathbb{D})$*
3) *for all $I \in \mathfrak{B}_{[0,1]}$, $\mathbb{P}(f^{-1}(I)) = \lambda(I)$*

We believe that the results we prove in the following not only improve our understanding of polar and Reed-Muller codes: Since its introduction in 2009, the polarization technique proposed by Arıkan has found its way into areas different from polar coding. Haghighatshoar and Abbe showed in the context of compression of analog sources that Rényi information dimension can be polarized [8], and Abbe and Wigderson used polarization for the construction of high-girth matrices [9]. Recently, Nasser proved that a binary operation is polarizing if and only if it is uniformity preserving and its inverse is strongly ergodic [10], [11]. We believe that our results might carry over to these areas as well; Section VI points to possible extensions.

Parts of this work have been presented at the 2015 NEWCOM# Emerging Topics in Modulation and Coding Workshop and will be presented at the 2016 International Zürich Seminar on Communications.

The work of Bernhard C. Geiger has been funded by the Erwin Schrödinger Fellowship J 3765 of the Austrian Science Fund and by the German Ministry of Education and Research in the framework of an Alexander von Humboldt Professorship.

Bernhard C. Geiger (geiger@ieee.org) is with the Institute for Communications Engineering, Technical University of Munich.

II. Preliminaries for Polar Codes

We adopt the notation of [2]. Let $W\colon \{0,1\} \to \mathcal{Y}$ be a binary-input memoryless channel with output alphabet $\mathcal{Y}$, capacity $0 < I(W) < 1$, and with Bhattacharyya parameter

$$Z(W) := \sum_{y\in\mathcal{Y}} \sqrt{W(y|0)W(y|1)}. \tag{3}$$

That $Z(W) = 0 \Leftrightarrow I(W) = 1$ and $Z(W) = 1 \Leftrightarrow I(W) = 0$ is a direct consequence of [2, Prop. 1]. We say a channel is *symmetric* if there exists a permutation $\pi\colon \mathcal{Y} \to \mathcal{Y}$ such that $\pi^{-1} = \pi$ and, for every $y \in \mathcal{Y}$, $W(y|0) = W(\pi(y)|1)$.

The heart of Arıkan's polarization technique is that two channel uses of W can be *combined and split* into one use of a "worse" channel

$$W_2^0(y_1^2|u_1) := \frac{1}{2}\sum_{u_2} W(y_1|u_1 \oplus u_2)W(y_2|u_2) \tag{4a}$$

and one use of a "better" channel

$$W_2^1(y_1^2, u_1|u_2) := \frac{1}{2} W(y_1|u_1 \oplus u_2)W(y_2|u_2) \tag{4b}$$

where $u_1, u_2 \in \{0,1\}$ and $y_1, y_2 \in \mathcal{Y}$. In essence, the combining operation codes two input bits by F in (1) and transmits the coded bits over W via two channel uses, creating a vector channel. The splitting operation splits this vector channel into the two virtual binary-input memoryless channels indicated in (4). Of these, the better (worse) channel has a strictly larger (smaller) capacity than the original channel W, i.e., $I(W_2^0) < I(W) < I(W_2^1)$, while the sum capacity equals twice the capacity of the original channel, i.e., $I(W_2^0) + I(W_2^1) = 2I(W)$ [2, Prop. 4].

The effect of combining and splitting on the channel capacities $I(W_2^0)$ and $I(W_2^1)$ admits no closed-form expression; the effect on the Bhattacharyya parameter at least admits bounds:

Lemma 2 ([2, Prop. 5 & 7])**.**

$$Z(W_2^1) = g_1(Z(W)) := Z^2(W) < Z(W) \tag{5a}$$

$$Z(W) < Z(W_2^0) \leq g_0(Z(W)) := 2Z(W) - Z^2(W) \tag{5b}$$

with equality if W is a binary erasure channel.

Channels with larger blocklengths 2^n, $n > 1$, can either be obtained by direct n-fold combining (using the matrix $G(n)$) and n-fold splitting, or by recursive pairwise combining and splitting. For $b^n \in \{0,1\}^n$, we obtain

$$\left(W_{2^n}^{b^n}, W_{2^n}^{b^n}\right) \to \left(W_{2^{n+1}}^{b^n0}, W_{2^{n+1}}^{b^n1}\right) \tag{6}$$

where b^n0 and b^n1 denote the sequences of zeros and ones obtained by appending 0 and 1 to b^n, respectively. Note that g_1 and g_0 from Lemma 2 are non-negative and non-decreasing functions mapping the unit interval onto itself, hence the inequality in (5b) is preserved under composition:

$$Z(W_{2^n}^{b^n}) \leq p_{b^n}(Z(W)) := g_{b_n}(g_{b_{n-1}}(\cdots g_{b_1}(Z(W))\cdots)) \tag{7}$$

The channel polarization theorem shows that, with probability one, after infinitely many combinations and splits, only perfect or useless channels remain, i.e., either $I(W_\infty^b) = 1$ or $I(W_\infty^b) = 0$ for $b \in \{0,1\}^\infty$. This is made precise in:

Proposition 1 ([2, Prop. 10])**.** *With probability one, the limit RV $I_\infty(b) := I(W_\infty^b)$ takes values in the set $\{0,1\}$: $\mathbb{P}(I_\infty = 1) = I(W)$ and $\mathbb{P}(I_\infty = 0) = 1 - I(W)$.*

This immediately gives rise to

Definition 1 (The Good and the Bad Channels)**.** Let $\mathcal{G}$ denote the set of good channels, i.e.,

$$x \in \mathcal{G} \Leftrightarrow \exists b \in f^{-1}(x)\colon I(W_\infty^b) = 1. \tag{8a}$$

Let $\mathcal{B}$ denote the set of bad channels, i.e.,

$$x \in \mathcal{B} \Leftrightarrow \exists b \in f^{-1}(x)\colon I(W_\infty^b) = 0. \tag{8b}$$

If the polarization procedure is stopped at a finite blocklength 2^n for n large enough, it can still be shown that the vast majority of the resulting 2^n channels are either almost perfect or almost useless, in the sense that the channel capacities are close to one or to zero (or that the corresponding Bhattacharyya parameters are close to zero or to one). The idea of polar coding is to transmit data only on those channels that are almost perfect: n-fold combining and splitting leads to 2^n virtual channels, each corresponding to a row of $G(n)$. The channels with high capacity are indicated by $\mathcal{F}$, and the generator matrix of the corresponding polar code is the submatrix of $G(n)$ consisting of those indicated rows. If the blocklength grows to infinity ($n \to \infty$), the set $\mathcal{F}$ becomes equivalent to the set $\mathcal{G}$ in Definition 1.

The difficulty of polar coding lies in code construction, i.e., in determining *which* channels/row indices are in the sets $\mathcal{F}$ and $\mathcal{G}$ for finite and infinite blocklengths. This immediately translates to the question which sequences $b \in \{0,1\}^\infty$ correspond to combinations and splits leading to a perfect channel (or which finite-length sequences b^n lead to channels with capacity sufficiently close to one). Determining the capacity of the virtual channels is an inherently difficult operation, since, whenever W is not a binary erasure channel (BEC), the cardinality of the output alphabet increases exponentially in 2^n [12, Ch. 3.3], [13, p. 36]. To circumvent this problem, Tal and Vardy presented an approximate construction method in [14], that relies on reduced output alphabet channels that are either upgraded or degraded w.r.t. the channel of interest. As these upgrading/degrading properties – mentioned earlier in Korada's PhD thesis [13, Def. 1.7 & Lem. 1.8] – play a fundamental role in this work, we present

Definition 2 (Channel Up- and Degrading)**.** A channel $W^-\colon \{0,1\} \to \mathcal{Z}$ is *degraded* w.r.t. the channel W (short: $W^- \preccurlyeq W$) if there exists a channel $Q\colon \mathcal{Y} \to \mathcal{Z}$ such that

$$W^-(z|u) = \sum_{y\in\mathcal{Y}} W(y|u)Q(z|y). \tag{9}$$

A channel $W^+\colon \{0,1\} \to \mathcal{Z}$ is *upgraded* w.r.t. the channel W (short: $W^+ \succcurlyeq W$) if there exists a channel $P\colon \mathcal{Z} \to \mathcal{Y}$ such that

$$W(y|u) = \sum_{z\in\mathcal{Z}} W^+(z|u)P(y|z). \tag{10}$$

Moreover, $W^+ \succcurlyeq W$ if and only if $W \preccurlyeq W^+$.

The upgraded (degraded) approximation remains upgraded (degraded) during combining and splitting:

Lemma 3 ([13, Lem. 4.7] & [14, Lem. 3])**.** *Assume that $W^- \preccurlyeq W \preccurlyeq W^+$. Then,*

$$
\begin{aligned}
I(W^-) &\le I(W) \le I(W^+) && \text{(11a)}\\
Z(W^-) &\ge Z(W) \ge Z(W^+) && \text{(11b)}\\
(W^-)_2^1 &\preccurlyeq \ \ W_2^1 \ \ \preccurlyeq (W^+)_2^1 && \text{(11c)}\\
(W^-)_2^0 &\preccurlyeq \ \ W_2^0 \ \ \preccurlyeq (W^+)_2^0. && \text{(11d)}
\end{aligned}
$$

It can be shown that the better channel (4b) obtained from combining and splitting is upgraded w.r.t. the original channel (as already mentioned in [12, p. 9]). The worse channel (4a) is degraded at least if W is symmetric.

Lemma 4 ([12, p. 9] & [6, Lem. 3])**.** *$W \preccurlyeq W_2^1$. If W is symmetric, then $W_2^0 \preccurlyeq W \preccurlyeq W_2^1$.*

Proof: By choosing

$$
P(y|y_1^2, u_1) = \begin{cases} 1, & \text{if } y = y_2 \\ 0, & \text{else.} \end{cases} \tag{12}
$$

one can show that $W \preccurlyeq W_2^1$. To show that also $W_2^0 \preccurlyeq W$ for symmetric channels, take [6, Lem. 3]

$$
Q(y_1^2|y) = \begin{cases} \frac{1}{2}W(y_2|0) & \text{if } y_1 = y \\ \frac{1}{2}W(y_2|1) & \text{if } y_1 = \pi(y) \\ 0 & \text{else} \end{cases} . \tag{13}
$$

■

Example 2. For a BEC W with erasure probability ϵ, W_2^1 is a BEC with erasure probability ϵ^2 and W_2^0 is a BEC with erasure probability $2\epsilon - \epsilon^2$ [2, Prop. 6]. The channel W_2^1 is an upgrade of W, because it can be degraded to W by appending a BEC with erasure probability $\epsilon/(1+\epsilon)$. The channel W_2^0 is degraded w.r.t. W by appending a BEC with erasure probability ϵ.

III. Properties of the Sets $\mathcal{G}$ and $\mathcal{B}$

In this section we develop the properties of the sets of good and bad channels.

Proposition 2. *For almost all x, there exists a value $0 \le \vartheta(x) \le 1$ such that $Z(W) < \vartheta(x)$ implies $x \in \mathcal{G}$. If W is a BEC, then additionally $Z(W) > \vartheta(x)$ implies $x \in \mathcal{B}$.*

Proof: See Appendix A. ■

If W is not a BEC, it may happen that $Z(W) > \vartheta(f(b))$ while still $I(W_\infty^b) = 1$. This leads to the question whether the set of good channels is (almost surely) increasing with decreasing Bhattacharyya parameter, i.e., if the sets of good channels for W and W' with $Z(W) > Z(W')$ are *aligned*. While in general the answer is negative [4], Proposition 2 answers it positively if W is a BEC: The set of good channels for a BEC is also good for any binary-input memoryless channel with a smaller Bhattacharyya parameter [15].

Example 3. For $x \in \mathbb{D}$, $\vartheta(x) = 1$: If $Z(W) < 1$, i.e., if the channel is not completely useless a priori, the non-terminating expansion of x will make it a perfect channel (cf. Proposition 4).

In Appendix B we prove that the thresholds of Proposition 2 are symmetric:

Proposition 3. *For those $x \notin \mathbb{D}$ for which $\vartheta(x)$ exists, $\vartheta(1-x) = 1 - \vartheta(x)$.*

The case $x \in \mathbb{Q} \setminus \mathbb{D}$ is interesting. In this case, the binary expansion is unique and *recurring*, i.e., there is a length-k sequence $a^k \in \{0,1\}^k$, such that $f(b^n a^k a^k a^k \cdots) = x$ for some $b^n \in \{0,1\}^n$. It is straightforward to show that for every non-trivial sequence a_k (i.e., a_k contains zeros and ones), p_{a^k} is from $[0,1]$ to $[0,1]$, non-negative, and non-decreasing, with vanishing derivatives at 0 and 1. Since this ensures that $p_{a^k}(z) < z$ for z close to zero and $p_{a^k}(z) > z$ for z close to one, the operation $z_{i+1} = p_{a^k}(z_i)$ constitutes an iterated function system with attracting fixed points at $z = 0$ and $z = 1$. Note further that, since p_{a^k} corresponds to the recurring part of the binary expansion of x, $Z(W_\infty^{b^n a^k a^k \cdots})$ will be bounded from above by the value to which this iterated function system converges after being initialized with $Z(W_{2^n}^{b^n})$. To show that Proposition 2 holds for $x \in \mathbb{Q} \setminus \mathbb{D}$ requires showing that p_{a^k} intersects the identity function only once on $(0,1)$, i.e., that there is no attracting fixed point on this open interval. We leave this problem for future investigation.

Example 4. Let $x = 2/3$, hence $f^{-1}(x) = 101010101\cdots$. It suffices to consider one period of the recurring sequence and determine its fixed points. In this case we get $p_{10}(z) = 2z^2 - z^4$. Its fixed points are the roots of $p_{10}(z) - z$; removing the trivial roots at $z = 0$ and $z = 1$ leaves two further roots at $(\pm\sqrt{5} - 1)/2$. One of these roots lies outside $[0,1]$ and is hence irrelevant. The remaining root determines the threshold, $\vartheta(2/3) = (\sqrt{5} - 1)/2$.

Let W be a BEC with erasure probability $\epsilon = Z(W) = \vartheta(2/3)$. Since $\epsilon = \vartheta(2/3)$ is a fixed point of the iterated function system corresponding to the recurring binary expansion, one gets $Z(W_\infty^{f^{-1}(2/3)}) = \epsilon \notin \{0,1\}$. This example illustrates that Proposition 1 holds only almost surely.

Proposition 4. *$\mathcal{G} \cap \mathcal{B} = \mathbb{D}$.*

Proof: See Appendix C. ■

That the intersection of the sets of good and bad channels is non-empty is a direct consequence of the non-injectivity of f. Note further that this intersection cannot be larger, since $\mathbb{D}$ is the only set to which f maps non-injectively. Since $\mathbb{D}$, a common subset of $\mathcal{G}$ and $\mathcal{B}$, is dense in $[0,1]$, both the set of good channels and the set of bad channels are dense in the unit interval. But even if dyadic rationals are excluded, results about denseness can be proved:

Proposition 5. *$\mathcal{G} \setminus \mathbb{D}$ is dense in $[0,1]$. If W is a BEC, then also $\mathcal{B} \setminus \mathbb{D}$ is dense in $[0,1]$.*

Proof: See Appendix D. ■

The proposition states that, at least for the BEC, there is no interval which contains only good channels. Hence, given a specific channel $W_{2^n}^{b^n}$, it is not possible to assume that a well-specified subset of channels (e.g., all $W_\infty^{b^n a}$ for a starting with 1) generated from this channel by combining and splitting will be perfect. The construction algorithm for an infinite-blocklength, *vanishing-error* polar code hence cannot stop at a finite blocklength. This is in contrast with finite-blocklength polar codes, for which an approximate construction technique

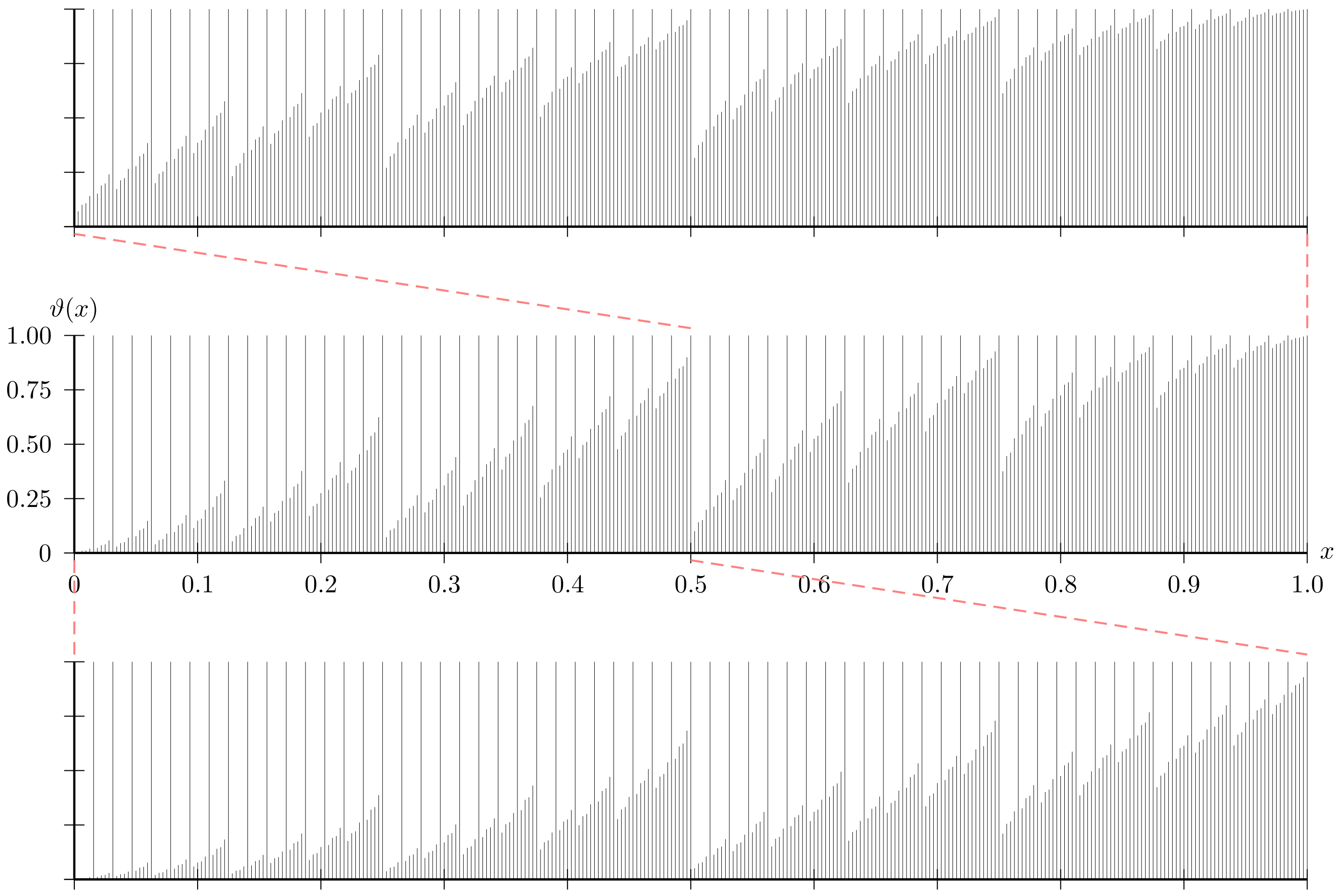

Fig. 1. The polar fractal for a BEC. The center plot shows the thresholds $\vartheta(x)$ for $x \in [0,1]$, while the bottom and the top plots show these thresholds for the scaled and shifted sets $[0,0.5]$ and $[0.5,1]$, respectively. Hence, the thresholds in the top plot are larger than the thresholds in the center plot, which are larger than those in the bottom plot. The indicator function of $\mathcal{G}$ is obtained by setting each value in the plot to one (zero) if the erasure probability ϵ is smaller (larger) than the threshold. Note further that the figure illustrates the symmetry of $\vartheta(x)$ mentioned in Proposition 3.

suggests to stop polarizing some channels at already shorter blocklengths [16].

Proposition 6. *$\mathcal{G}$ is Lebesgue measurable and has Lebesgue measure $\lambda(\mathcal{G}) = I(W)$. $\mathcal{B}$ is Lebesgue measurable and has Lebesgue measure $\lambda(\mathcal{B}) = 1 - I(W)$.*

Proof: See Appendix E. ■

Note that $\lambda(\mathcal{G} \cup \mathcal{B}) = 1$ although $\mathcal{G} \cup \mathcal{B} \subset [0,1]$. The reason is that convergence to good or bad channels is only almost sure, i.e., there may be channels W_∞^b that are neither good nor bad (see Example 4).

An immediate consequence of Proposition 6 is that $\mathcal{G}$ and $\mathcal{B}$ have a Hausdorff dimension equal to one. This follows from the fact that the one-dimensional Hausdorff measure of a set equals its Lebesgue measure up to a constant [17, eq. (3.4), p. 45]. Since, thus, the one-dimensional Hausdorff measures of $\mathcal{G}$ and $\mathcal{B}$ are positive and finite, we have

Corollary 1. *The Hausdorff dimensions of $\mathcal{G}$ and $\mathcal{B}$ satisfy $d(\mathcal{G}) = 1$ and $d(\mathcal{B}) = 1$.*

Also the box-counting dimensions [17, p. 28] are equal to one, since both sets are dense on the unit interval [17, Prop. 2.6].

We finally come to the claim that polar codes are fractal. Following Falconer's definition [17, p. xxviii], a set is fractal if it is (at least approximately) self-similar and has detail on arbitrarily small scales, or if its fractal dimension (e.g., its Hausdorff dimension) is larger than its topological dimension. Whether or not the result shown below will convince the reader of this property is a mere question of definition; strictly speaking, we can show only *quasi self-similarity* of $\mathcal{G}$:

Proposition 7. *Let $\mathcal{G}_n(k) := \mathcal{G} \cap [(k-1)2^{-n}, k2^{-n}]$ for $k = 1, \dots, 2^n$. $\mathcal{G} = \mathcal{G}_0(1)$ is* quasi self-similar *in the sense that, for all n and all k, $\mathcal{G}_n(k) = \mathcal{G}_{n+1}(2k-1) \cup \mathcal{G}_{n+1}(2k)$ is quasi self-similar to its right half:*

$$\mathcal{G}_n(k) \subset 2\mathcal{G}_{n+1}(2k) - k2^{-n} \tag{14}$$

If W is symmetric, $\mathcal{G}_n(k)$ is quasi self-similar:

$$2\mathcal{G}_{n+1}(2k-1) - (k-1)2^{-n} \subset \mathcal{G}_n(k) \subset 2\mathcal{G}_{n+1}(2k) - k2^{-n} \tag{15}$$

Proof: See Appendix F. ■

In other words, at least for a symmetric channel, $\mathcal{G}$ is composed of two similar copies of itself (see Fig. 1). The self-similarity is closely related to the fact that polar codes are decreasing monomial codes [6, Thm. 1]. Along the same lines, the quasi self-similarity of $\mathcal{B}$ can be shown.

Example 5. By careful computations we obtain $\vartheta(1/6) \approx 0.214$, $\vartheta(1/3) \approx 0.382$, and $\vartheta(2/3) \approx 0.618$. Indeed, if we consider $1/3$ in $\mathcal{G}$, then $1/6$ and $2/3$ are the corresponding values in $\mathcal{G}_1(1)$ and $\mathcal{G}_1(2)$. Since $\vartheta(1/6) < \vartheta(1/3) < \vartheta(2/3)$, for the BEC we have the inclusion indicated in Proposition 7.

IV. Preliminaries for Reed-Muller Codes

As mentioned above, a rate-$K/2^n$ Reed-Muller code has a $K \times 2^n$ generator matrix with all K rows having a Hamming weight larger than a predefined threshold. To make this more precise, let $w(b^n) = \sum_{i=1}^n b_i$ be the *Hamming weight* of $b^n \in \{0,1\}^n$ and let $s_i(n)$ be the i-th row of $G(n)$. The generator matrix $G_{RM}(r,n)$ of an order-r, length-2^n Reed-Muller code consists of the rows of $G(n)$ indicated in [3]

$$\mathcal{F} = \{i \in \{1,\dots,2^n\}\colon w(s_i(n)) \geq 2^{n-r}\}. \tag{16}$$

Trivially, $G_{RM}(n,n) = G(n)$, while $G_{RM}(0,n)$ is a single row vector containing only ones (length-2^n repetition code).

To analyze the effect of doubling the block length, note that

$$G(n+1) := \begin{bmatrix} G(n) & 0 \\ G(n) & G(n) \end{bmatrix}. \tag{17}$$

Assume that we indicate the rows of $G(n)$ by a sequence of binary numbers, i.e., let the i-th row be indexed by $h_n(b^n) := 2^n \sum_{l=1}^n b_l 2^{-l}$. Furthermore, let $0b^n$ and $1b^n$ denote the sequences of zeros and ones obtained be prepending 0 and 1 to b^n, respectively. Clearly, $h_{n+1}(0b^n) = h_n(b^n)$ and $h_{n+1}(1b^n) = h_n(b^n) + 2^n$. Combining this with (17) yields

$$w(s_{h_{n+1}(0b^n)}(n+1)) = w(s_{h_n(b^n)}(n)) \tag{18}$$

$$w(s_{h_{n+1}(1b^n)}(n+1)) = 2w(s_{h_n(b^n)}(n)). \tag{19}$$

Defining $G(0) := 1$, we thus get

$$w(s_{h_n(b^n)}(n)) = 2^{w(b^n)} \tag{20}$$

and

$$\mathcal{F} = h_n\left(\{b^n \in \{0,1\}^n\colon 2^{w(b^n)} \geq 2^{n-r}\}\right). \tag{21}$$

Letting the blocklengths go to infinity, we may ask questions about the following set:

Definition 3 (The Heavy Channels)**.** Let $\mathcal{H}(\rho)$ denote the set of ρ-heavy channels, i.e.,

$$x \in \mathcal{H}(\rho) \Leftrightarrow \exists b \in f^{-1}(x)\colon \liminf_{n\to\infty} \frac{2^{w(b^n)}}{2^{n\rho}} \geq 1. \tag{22}$$

Loosely speaking, the set of heavy channels corresponds to those rows of $G(n)$, which *asymptotically* have a Hamming weight larger than a given threshold.

Example 6. $\mathcal{H}(1) = \{1\}$. This follows from the fact that 1 is the only number in the unit interval with a binary expansion consisting only of ones. $\mathcal{H}(0) = [0,1]$. This follows from the fact that $w(b^n) \geq 0$.

The results we will show for the set $\mathcal{H}(\rho)$ are tightly linked to the concept of *normal numbers.*

Definition 4 (Normal Numbers)**.** A number $x \in [0,1]$ is called *simply normal to base 2* ($x \in \mathcal{N}$) iff

$$\exists b \in f^{-1}(x)\colon \lim_{n\to\infty} \frac{w(b^n)}{n} = \frac{1}{2}. \tag{23}$$

In general, a number is simply normal in base M if the number of each of its digits used in its M-ary expansion is $1/M$. A number is called normal if this property not only holds for digits, but for subsequences: a number is normal in base M if, for each $k \geq 1$, the number of each of its length-k sequences used in its M-ary expansion is $1/M^k$. It immediately follows that a normal number is simply normal. The converse is in general not true:

Example 7. Let $x = 1/3$, hence $b = 010101\cdots$. x is simply normal to base 2, but not normal (since the sequences 00 and 11 never occur). Let $x = 1/7$, hence $b = 001001001\cdots$. x is neither normal nor simply normal. Let $x \in \mathbb{D}$, hence b is either terminating ($\lim_{n\to\infty} w(b^n)/n = 0$) or non-terminating ($\lim_{n\to\infty} w(b^n)/n = 1$). Dyadic rationals are not simply normal.

Lemma 5 (Borel's Law of Large Numbers, cf. [18, Cor. 8.1, p. 70])**.** *Almost all numbers in* $[0,1]$ *are simply normal, i.e.,*

$$\lambda(\mathcal{N}) = 1. \tag{24}$$

Although normal numbers are, in this sense, *normal,* there are uncountably many numbers in the unit interval which are not normal. Moreover, the set of numbers that are not normal is *superfractal,* i.e., it has a Hausdorff dimension equal to one although it has zero Lebesgue measure [19].

V. Properties of the Set $\mathcal{H}$

We can show in Appendix G that the dyadic rationals are not only good and bad, but also heavy:

Proposition 8. *For all* $\rho \in [0,1)$, $\mathbb{D} \subset \mathcal{H}(\rho)$.

It follows that $\mathcal{H}(\rho)$ is dense in $[0,1]$ for all $\rho \in [0,1)$.

The Lebesgue measure of the set of good channels was equal to the channel capacity of W. The result for heavy channels is inherently different, because $\mathcal{H}(\rho)$ does not depend on W. The proof of the following result can be found in Appendix H.

Proposition 9. $\mathcal{H}(\rho)$ *is Lebesgue measurable and has Lebesgue measure*

$$\lambda(\mathcal{H}(\rho)) = \begin{cases} 1, & \textit{if } \rho < 1/2 \\ 0, & \textit{if } \rho \geq 1/2 \end{cases}. \tag{25}$$

The result is surprising since it suggests a *phase transition* for the rate of Reed-Muller codes: If $\rho < 1/2$, the infinite-blocklength Reed-Muller code consists of almost all (in the sense of Lebesgue measure) possible binary sequences. In contrast, if $\rho \geq 1/2$, the infinite-blocklength Reed-Muller code consists of almost no code words (again, in the sense of Lebesgue measure). The picture is not as simple if one also considers the Hausdorff dimension of $\mathcal{H}(\rho)$. In Appendix I we prove that $\mathcal{H}(\rho)$ has positive Hausdorff dimension even if it is a Lebesgue null set.

Proposition 10. *The Hausdorff dimension satisfies*

$$d(\mathcal{H}(\rho)) \begin{cases} = 1, & \text{if } \rho \le 1/2 \\ \ge h_2(\rho), & \text{if } \rho > 1/2 \end{cases} \tag{26}$$

where $h_2(x) := -x\log_2 x - (1-x)\log_2(1-x)$.

Unfortunately, we were not able to give an exact expression for the Hausdorff dimension of $\mathcal{H}(\rho)$ for $\rho > 1/2$. While the set of all non-normal numbers is superfractal, we are not sure if this holds also for a proper subset.

The sets $\mathcal{G}$ and $\mathcal{B}$ exhibit self-similarity, i.e., detailed structure on every scale (cf. Fig. 1). We next show that also $\mathcal{H}(\rho)$ is self-similar. At least for $\mathcal{H}(0)$ and $\mathcal{H}(1)$ (cf. Example 6) this is as trivial as the self-similarity of a point or a line. For $\rho \in (0,1)$ this self-similarity is more interesting, and related to the fact that Reed-Muller codes are decreasing monomial codes [6, Prop. 2]. In Appendix J we prove

Proposition 11. *Let* $\mathcal{H}_n(\rho,k) := \mathcal{H}(\rho) \cap [(k-1)2^{-n}, k2^{-n}]$ *for* $k = 1,\dots,2^n$. $\mathcal{H}(\rho) = \mathcal{H}_0(\rho,1)$ *is* quasi self-similar *in the sense that, for all* n *and all* k, $\mathcal{H}_n(\rho,k) = \mathcal{H}_{n+1}(\rho,2k-1) \cup \mathcal{H}_{n+1}(\rho,2k)$ *is quasi self-similar:*

$$\begin{aligned} &2\mathcal{H}_{n+1}(\rho,2k-1) - (k-1)2^{-n} \subset \\ &\qquad \mathcal{H}_n(\rho,k) \subset 2\mathcal{H}_{n+1}(\rho,2k) - k2^{-n}. \end{aligned} \tag{27}$$

VI. Discussion & Outlook

That polar codes satisfy fractal properties has long been suspected: Every nontrivial, partly polarized channel $W_{2^n}^{b^n}$ gives rise, by further polarization, to both perfect and useless channels, regardless how close $I(W_{2^n}^{b^n})$ is to zero or one. This fact is reflected in our Propositions 4 and 5, which state that the good channels are dense in the unit interval (and so are the bad channels for BECs): A partial polarization with sequence b^n corresponds to an interval with dyadic endpoints, and denseness implies that in this interval there will be both perfect and useless channels. Proposition 7, claiming the self-similarity of the sets of good and bad channels, goes one step further and gives these sets structure: If a channel polarized according to the sequence $b^n a$ is good, then so is the channel polarized according to $b^n 1a$.

An obvious extension of our work should deal with the fractal properties of non-binary polar and Reed-Muller codes. For example, if q is a prime number, then every invertible $\ell\times\ell$ matrix with entries from $\{0,\dots,q-1\}$ is polarizing, unless it is upper-triangular [12, Thm. 5.2]. The n-fold Kronecker product of one of these matrices generates ℓ^n channels. It is easy to design a function mapping $\{0,\dots,\ell-1\}^\infty$ to $[0,1]$ (cf. (2)), admitting an analysis similar to the one presented in this paper. Along the same lines, it would be interesting to examine the properties of q-ary Reed-Muller codes, e.g., [20], [21].

Whether binary or not, it is presently not clear how our infinite-blocklength results can be carried over to practically relevant finite-length codes. Future work shall investigate this issue.

Acknowledgments

The author thanks Emmanuel Abbe, Princeton University, and Hamed Hassani, ETH Zurich, for fruitful discussions and suggesting material. The author is particularly indebted to Jean-Pierre Tillich, INRIA, for helpful suggestions and generalizing Proposition 7.

Appendix A
Proof of Proposition 2

Recall that, by Lemma 2, we have

$$Z(W_{2^n}^{b^n}) \le p_{b^n}(Z(W)) := g_{b_n}(g_{b_{n-1}}(\cdots g_{b_1}(Z(W))\cdots)).$$

Lemma 6 ([22, Lem. 11])**.** *For* $\mathbb{P}$*-almost every realization* $b \in \Omega$, *there exists a point* $\theta(b) \in [0,1]$, *such that*

$$\lim_{n\to\infty} p_{b^n}(z) = \begin{cases} 0, & z \in [0,\theta(b)) \\ 1, & z \in (\theta(b),1] \end{cases}. \tag{28}$$

Furthermore, the thus constructed RV θ *is uniformly distributed on* $[0,1]$.

If $Z(W) < \theta(b)$, $Z(W_\infty^b) \le \lim_{n\to\infty} p_{b^n}(Z(W)) = 0$, and hence $f(b) \in \mathcal{G}$. We now define $\vartheta(f(b)) := \theta(b)$ if $f(b) \notin \mathbb{D}$ and $\vartheta(f(b)) = 1$ if $f(b) \in \mathbb{D}$ (since $\mathbb{D} \subset \mathcal{G}$ by Proposition 4).

Proof for BECs: If W is a BEC, then $Z(W_{2^n}^{b^n}) = p_{b^n}(Z(W))$. Hence, by Lemma 6, if $\epsilon < \theta(b)$, then $Z(W_\infty^b) = \lim_{n\to\infty} p_{b^n}(\epsilon) = 0$, and if $\epsilon > \theta(b)$, then $Z(W_\infty^b) = \lim_{n\to\infty} p_{b^n}(\epsilon) = 1$. ■

Appendix B
Proof of Proposition 3

Let $b \in f^{-1}([0,1]\setminus\mathbb{D}) \subset \{0,1\}^\infty$, and let $\bar{b}$ be such that $\bar{b}_i = 1 - b_i$ for all i. It follows from the linearity of f that $f(b) + f(\bar{b}) = f(b+\bar{b}) = 1$, because $b + \bar{b} = 11111\cdots$. Hence, if $x \notin \mathbb{D}$ has binary expansion b, then $1-x$ has binary expansion $\bar{b}$. It can be easily verified that $g_i(1-z) = 1 - g_{1-i}(z)$ for $i = 0,1$. Hence,

$$\begin{aligned} p_{b^n}(z) &= g_{b_n}(g_{b_{n-1}}(\cdots g_{b_2}(g_{b_1}(z))\cdots)) \\ &= g_{b_n}(g_{b_{n-1}}(\cdots g_{b_2}(1-g_{\bar{b}_1}(1-z))\cdots)) \\ &= g_{b_n}(g_{b_{n-1}}(\cdots 1-g_{\bar{b}_2}(g_{\bar{b}_1}(1-z))\cdots)) \\ &= 1-g_{\bar{b}_n}(g_{\bar{b}_{n-1}}(\cdots g_{\bar{b}_2}(g_{\bar{b}_1}(1-z))\cdots)) \\ &= 1-p_{\bar{b}^n}(1-z). \end{aligned}$$

If $0 \le z < \theta(b)$, then $1-\theta(b) < 1-z \le 1$. Since $0 \le z < \theta(b)$ implies $p_{b^n}(z) \to 0$ and $p_{\bar{b}^n}(1-z) \to 1$, we get $\theta(\bar{b}) = 1-\theta(b)$, and hence $\vartheta(1-x) = 1-\vartheta(x)$. This completes the proof. ■

Appendix C
Proof of Proposition 4

That $\mathcal{G} \cap \mathcal{B} \subseteq \mathbb{D}$ follows from the fact that only dyadic rationals have a non-unique binary expansion. In particular, the preimage of every $x \in \mathbb{D}$ consists of two elements, namely

$$(b^{n-1}b_n 0000000\cdots) \tag{29a}$$

and

$$(b^{n-1}\bar{b}_n 1111111\cdots) \tag{29b}$$

where $\overline{b}_n = 1 - b_n$. By the properties of combining and splitting,

$$0 < I(W_{2^n}^{b^{n-1}\overline{b}_n}), I(W_{2^n}^{b^{n-1}b_n}) < 1. \tag{30}$$

We first show that (29b) leads to a good channel. To this end, observe that, by [2, Prop. 7], the Bhattacharyya parameter satisfies $0 < Z(W_{2^{n+1}}^{b^{n-1}\overline{b}_n 1}) = Z(W_{2^n}^{b^{n-1}\overline{b}_n})^2 < 1$. Iterating the squaring operation drives the Bhattacharyya parameter to zero, i.e., $Z(W_\infty^{b^{n-1}\overline{b}_n 1111111\cdots}) = 0$, hence $I(W_\infty^{b^{n-1}\overline{b}_n 1111111\cdots}) = 1$ and $\mathbb{D} \in \mathcal{G}$.

To show that (29a) leads to a bad channel, assume that $I(W_\infty^{b^n 0000000\cdots}) = \delta$. We now show that for every $a \in \Omega$, $I(W_\infty^{b^n 0000000\cdots}) \leq I(W_\infty^{b^n a})$. For example, take $a = (1011100\cdots)$. By Lemmas 3 (a) and 4 (b), the following list of relations can be shown:

$$\begin{aligned}
W_\infty^{b^n} &\overset{(a)}{\preccurlyeq} W_\infty^{b^n 1}\\
W_\infty^{b^n 0} &\overset{(b)}{\preccurlyeq} W_\infty^{b^n 10}\\
W_\infty^{b^n 0} &\overset{(a)}{\preccurlyeq} W_\infty^{b^n 101}\\
W_\infty^{b^n 0} &\overset{(a)}{\preccurlyeq} W_\infty^{b^n 1011}\\
W_\infty^{b^n 0} &\overset{(a)}{\preccurlyeq} W_\infty^{b^n 10111}\\
W_\infty^{b^n 00} &\overset{(b)}{\preccurlyeq} W_\infty^{b^n 101110}\\
W_\infty^{b^n 000} &\overset{(b)}{\preccurlyeq} W_\infty^{b^n 1011100}\\
\ldots &\preccurlyeq \ldots
\end{aligned}$$

and hence, $W_\infty^{b^n 0000000\cdots} \preccurlyeq W_\infty^{b^n a}$. By Lemma 3, $\delta = I(W_\infty^{b^n 0000000\cdots}) \leq I(W_\infty^{b^n a})$ for every $a \in \Omega$, hence also

$$\delta \leq \inf_{a \in \Omega} I(W_\infty^{b^n a}). \tag{31}$$

But since $0 < I(W_{2^n}^{b^n}) < 1$, by Proposition 1 there must be sequences a such that $I(W_\infty^{b^n a}) = 0$, hence $\delta = 0$ and $\mathbb{D} \in \mathcal{B}$. ■

Appendix D
Proof of Proposition 5

The proof follows from showing that between every dyadic rational we can find a rational $x \in \mathbb{Q} \setminus \mathbb{D}$ such that $x \in \mathcal{G}$. To this end, fix $x_1 = p/2^n$ and $x_2 = (p+1)/2^n$. Let further b^n be the terminating binary expansion of x_1, i.e., $f(b^n 000\cdots) = x_1$. Let a^k be such that $a_1 = \cdots = a_{k-1} = 1$ and $a_k = 0$. Note that $x := f(b^n a^k a^k a^k \cdots) \in (x_1, x_2)$. We now bound the polynomial p_{a^k} from above:

$$p_{a^k}(z) = 2z^{2^{k-1}} - z^{2^k} \leq 2z^{2^{k-1}}$$

The bound crosses z at $z = 0$ and at $z^* = 2^{-1/(2^{k-1}-1)}$. From this follows that $p_{a^k}(z) < z$ for $z < z^*$, where z^* can be made arbitrarily close to one for k sufficiently large. Hence, if $z_{i+1} = p_{a^k}(z_i)$, then $z_i \to 0$ if $z_0 < z^*$. Let $z_0 = Z(W_{2^n}^{b_n})$ and let k be sufficiently large such that $z^* > z_0$. Then, $Z(W_\infty^{b^n a^k a^k \cdots}) = 0$ and $x \in \mathcal{G}$.

Proof for BECs: It remains to show that also $\mathcal{B} \setminus \mathbb{D}$ is dense in $[0, 1]$. To this end, we consider the sequence a^k such that $a_1 = \cdots = a_{k-1} = 0$ and $a_k = 1$. We now bound the polynomial p_{a^k} from below:

$$\begin{aligned}
p_{a^k}(z) &= \left(1 - (1-z)^{2^{k-1}}\right)^2\\
&= 1 - 2(1-z)^{2^{k-1}} + (1-z)^{2^{2k-2}}\\
&\geq 1 - 2(1-z)^{2^{k-1}}
\end{aligned}$$

The bound crosses z at at $z = 1$ and $z^* = 1 - 2^{-1/(2^{k-1}-1)}$. From this follows that $p_{a^k}(z) > z$ for $z > z^*$, where z^* can be made arbitrarily close to zero for k sufficiently large. Hence, if $z_{i+1} = p_{a^k}(z_i)$, then $z_i \to 1$ if $z_0 > z^*$. Let $z_0 = Z(W_{2^n}^{b_n})$ and let k be sufficiently large such that $z^* < z_0$. Then, $Z(W_\infty^{b^n a^k a^k \cdots}) = 1$ and $x \in \mathcal{B}$. ■

Appendix E
Proof of Proposition 6

In the proof we use Lemma 1. Note that $\lambda(\mathcal{G}) = \lambda(\mathcal{G} \setminus \mathbb{D}) + \lambda(\mathbb{D}) = \lambda(\mathcal{G} \setminus \mathbb{D})$, and similarly, $\lambda(\mathcal{B}) = \lambda(\mathcal{B} \setminus \mathbb{D})$. Since f is bijective on $\Omega' := \Omega \setminus f^{-1}(\mathbb{D})$, Definition 1 implies

$$x \in \mathcal{G} \setminus \mathbb{D} \Leftrightarrow I(W_\infty^{f^{-1}(x)}) = 1 \tag{32}$$

or $f^{-1}(\mathcal{G} \setminus \mathbb{D}) = \{b \in \Omega' \colon I(W_\infty^b) = 1\}$. Note further that $\mathbb{P}(f^{-1}(\mathbb{D})) = 0$. Hence, by Proposition 1,

$$\begin{aligned}
\lambda(\mathcal{G}) = \lambda(\mathcal{G} \setminus \mathbb{D}) &= \mathbb{P}(\{b \in \Omega' \colon I(W_\infty^b) = 1\})\\
&= \mathbb{P}(I_\infty = 1) = I(W).
\end{aligned} \tag{33}$$

The proof for the set of bad channels follows along the same lines. ■

Appendix F
Proof of Proposition 7

Since the dyadic rationals are self-similar and since, by Proposition 4, $\mathbb{D} \subset \mathcal{G}$, one has, for all n and k,

$$\mathcal{G}_n(k) \cap \mathbb{D} = 2\left(\mathcal{G}_{n+1}(2k) \cap \mathbb{D}\right) - k2^{-n}. \tag{34}$$

We now treat those values in $[0, 1]$ that are not dyadic rationals. If $b_k^n = b_1 b_2 \cdots b_n$ is the terminating binary expansion of $(k-1)2^{-n}$, every value in $[(k-1)2^{-n}, k2^{-n}]$ has a binary expansion $b_k^n a$ for some $a \in \Omega$, where $b_n = 1$ if and only if $(k-1)$ is odd. Similarly, and since $(2k-1)$ is always odd, every value in $[(2k-1)2^{-n-1}, k2^{-n}]$ has a binary expansion $b_k^n 1a'$ for some $a' \in \Omega$. Assume that $a' = a$. Then, by Lemmas 3 and 4, $W_\infty^{b_k^n a} \preccurlyeq W_\infty^{b_k^n 1a}$ for all a. Hence, if $f(b_k^n a) \in \mathcal{G}_n(k)$, then $f(b_k^n 1a) \in \mathcal{G}_{n+1}(2k)$. It remains to show that $2f(b_k^n 1a) - f(b_{k+1}^n) = f(b_k^n a)$:

$$\begin{aligned}
f(b_k^n a) + f(b_{k+1}^n) &= f(b_k^n) + 2^{-n} f(a) + f(b_{k+1}^n)\\
&= (k-1)2^{-n} + 2^{-n} f(a) + k2^{-n}\\
&= (2k-1)2^{-n} + 2^{-n} f(a)\\
&= 2(2k-1)2^{-n-1} + 2 \cdot 2^{-n-1} f(a)\\
&= 2f(b_k^n 1) + 2 \cdot 2^{-n-1} f(a)\\
&= 2f(b_k^n 1a)
\end{aligned}$$

Proof for Symmetric Channels: Since $(2k-2)$ is always even, every value in $[(2k-2)2^{-n-1}, (2k-1)2^{-n-1}]$

has a binary expansion $b_k^n 0a$ for some $a \in \Omega$. Then, by Lemmas 3 and 4, $W_\infty^{b_k^n 0a} \preccurlyeq W_\infty^{b_k^n a}$ for all a. Hence, if $f(b_k^n 0a) \in \mathcal{G}_{n+1}(2k)$, then $f(b_k^n a) \in \mathcal{G}_n(k)$. It remains to show that $2f(b_k^n 0a) - f(b_k^n) = f(b_k^n a)$:

$$\begin{aligned}
f(b_k^n a) + f(b_k^n) &= f(b_k^n) + 2^{-n} f(a) + f(b_k^n) \\
&= (k-1)2^{-n} + 2^{-n} f(a) + (k-1)2^{-n} \\
&= (2k-2)2^{-n} + 2^{-n} f(a) \\
&= 2(2k-2)2^{-n-1} + 2 \cdot 2^{-n-1} f(a) \\
&= 2f(b_k^n 0) + 2 \cdot 2^{-n-1} f(a) \\
&= 2f(b_k^n 0a)
\end{aligned}$$

■

Appendix G
Proof of Proposition 8

We take the non-terminating expansion of $x \in \mathbb{D}$, i.e., there is a $b^k \in \{0,1\}^k$ such that $f(b^k 1111\cdots) = x$. Hence, $w(b^n) \geq n - k$ for $n \geq k$. In Definition 3 we can take the binary logarithm on both sides of the inequality to get the condition

$$x \in \mathcal{H}(\rho) \Leftrightarrow \exists b \in f^{-1}(x)\colon \liminf_{n\to\infty} w(b^n) - n\rho \geq 0. \tag{35}$$

But

$$\liminf_{n\to\infty} w(b^n) - n\rho = \lim_{n\to\infty} n(1-\rho) - k \tag{36}$$

goes to infinity for $\rho < 1$. ■

Appendix H
Proof of Proposition 9

By Example 7, dyadic rationals are not simply normal, hence let $\mathcal{N} \subset [0,1]\backslash\mathbb{D}$ be the set of simply normal numbers in $[0,1]$. Note that f is bijective on $\mathcal{N}$ by Lemma 1. By Lemma 5 we have

$$\forall b \in f^{-1}(\mathcal{N})\colon w(b^n) = \frac{1}{2}n + o(n). \tag{37}$$

Fix ρ. Then,

$$\liminf_{n\to\infty} w(b^n) - n\rho = \lim_{n\to\infty} n\left(\frac{1}{2} - \rho\right) + o(n). \tag{38}$$

If $\rho < 1/2$, then this limit diverges to infinity, and hence $\mathcal{N} \subset \mathcal{H}(\rho)$. Thus, since $\lambda(\mathcal{N}) = 1$, we have $\lambda(\mathcal{H}(\rho)) = 1$. If $\rho > 1/2$, the limit diverges to minus infinity, and hence $\mathcal{N} \not\subset \mathcal{H}(\rho)$. Thus, $\mathcal{H}(\rho) \subset [0,1] \setminus \mathcal{N}$, from which follows $\lambda(\mathcal{H}(\rho)) = 0$.

Now let $\rho = 1/2$. We define a random variable B on our probability space, such that for all $b \in \Omega$, $B(b) = b$. B is a sequence of independent, identically distributed Bernoulli-1/2 random variables, i.e., for all i we have $\mathbb{P}(B_i = 1) = \mathbb{P}(B_i = 0) = 1/2$. We have

$$\lambda(\mathcal{H}(1/2)) = \mathbb{P}\left(\liminf_{n\to\infty} w(B^n) - \frac{n}{2} \geq 0\right). \tag{39}$$

Consider the simple random walk $S_n := w(B^n) - \frac{n}{2}$. Let $N_0(n)$ be the number of zero crossings of the sequence $S_1, \dots, S_n$, and let $N_0(n,b)$ be the number of zero crossings corresponding to the realization $b \in \Omega$. The event $\liminf_{n\to\infty} w(b^n) - \frac{n}{2} \geq 0$ can only happen if the realization of S_n corresponding to b has only finitely many zero crossings, i.e.,

$$\begin{aligned}
&\{b \in \Omega\colon \liminf_{n\to\infty} w(b^n) - \frac{n}{2} \geq 0\} \\
&\subseteq \{b \in \Omega\colon \exists R \in \mathbb{N}_0\colon \lim_{n\to\infty} N_0(n,b) \leq R\} \\
&= \bigcup_{R=0}^{\infty} \{b \in \Omega\colon \lim_{n\to\infty} N_0(n,b) \leq R\} \\
&= \bigcup_{R=0}^{\infty} \liminf_{n\to\infty} \{b \in \Omega\colon N_0(n,b) \leq R\}
\end{aligned}$$

and hence

$$\begin{aligned}
\mathbb{P}&\left(\{b \in \Omega\colon \liminf_{n\to\infty} w(b^n) - \frac{n}{2} \geq 0\}\right) \\
&\leq \sum_{R=0}^{\infty} \mathbb{P}(\liminf_{n\to\infty} \{b \in \Omega\colon N_0(n,b) \leq R\}) \\
&\leq \sum_{R=0}^{\infty} \lim_{n\to\infty} \mathbb{P}(N_0(n) \leq R)
\end{aligned} \tag{40}$$

where the second inequality is due to Fatou's lemma [23, Lem. 1.28, p. 23].

With [24, Ch. III.5, p. 84]

$$\mathbb{P}(N_0(n) = R) = 2\mathbb{P}(S_{2n+1} = 2R+1) \tag{41}$$

we get

$$\begin{aligned}
\mathbb{P}(N_0(n) \leq R) &= 2\sum_{r=0}^{R} \mathbb{P}(S_{2n+1} = 2r+1) \\
&\overset{(a)}{=} 2\sum_{r=0}^{R} \binom{2n+1}{n-r} 2^{-2n-1} \\
&\leq 2^{-2n} \sum_{r=0}^{R} \binom{2n+2}{n+1} \\
&= 2^{-2n}(R+1)\binom{2n+2}{n+1} \\
&\overset{(b)}{\leq} 2^{-2n}(R+1)e2^{2n+2}\frac{1}{\sqrt{(n+1)\pi}} \\
&= \frac{4e(R+1)}{\sqrt{(n+1)\pi}}
\end{aligned}$$

where (a) is [24, eq. (2.2), p. 75] and (b) is due to Stirling's approximation [25, eq. 6.1.38, p. 257]. Since for $n \to \infty$ this probability is zero, we have by (40)

$$\lambda(\mathcal{H}(1/2)) = \mathbb{P}\left(\liminf_{n\to\infty} w(B^n) - \frac{n}{2} \geq 0\right) = 0. \tag{42}$$

This completes the proof. ■

Appendix I
Proof of Proposition 10

That $d(\mathcal{H}(\rho)) = 1$ for $\rho < 1/2$ follows from Proposition 9 in combination with Corollary 1. For $\rho \geq 1/2$, we define

$$\tilde{\mathcal{N}_p} := \left\{x \in [0,1]\colon \exists b \in f^{-1}(x)\colon \lim_{n\to\infty} \frac{w(b^n)}{n} = p\right\} \tag{43}$$

for some $p \in (0,1)$. Note that $\tilde{\mathcal{N}}_{1/2} = \mathcal{N}$. By [26] (cf. [18, Chapter 8] for further notes), the Hausdorff dimension of this set is given by[1]

$$d(\tilde{\mathcal{N}}_p) = h_2(p). \tag{44}$$

Reasoning as in the proof of Proposition 9, $\tilde{\mathcal{N}}_p \subset \mathcal{H}(\rho)$ if $p > \rho$ and $\tilde{\mathcal{N}}_p \not\subset \mathcal{H}(\rho)$ if $p < \rho$. As a consequence,

$$\bigcup_{n=1}^{\infty} \tilde{\mathcal{N}}_{\rho+1/n} \subset \mathcal{H}(\rho). \tag{45}$$

For a countable sequence of sets A_n, Hausdorff dimension satisfies [17, p. 49]

$$d(\bigcup_{n=1}^{\infty} A_n) = \sup_{n\geq 1} d(A_n), \tag{46}$$

and hence, by the monotonicity of Hausdorff dimension [17, p. 48],

$$d(\mathcal{H}(\rho)) \geq \sup_{n\geq 1} h_2(\rho + 1/n) = h_2(\rho) \tag{47}$$

where the last equality follows from the fact that the binary entropy function decreases with increasing ρ for $\rho \geq 1/2$. In particular, for $\rho = 1/2$, $d(\mathcal{H}(\rho)) = 1$. This completes the proof. ■

Appendix J
Proof of Proposition 11

The proof follows along the lines of the proof of Proposition 7. Let again b_k^n be the terminating expansion of $(k-1)2^{-n}$, and let $a \in \Omega$. The connections between the sequences $b := b_k^n a$, $b_- := b_k^n 0a$, and $b_+ := b_k^n 1a$ have been established above. To prove the theorem, we have to show that

$$\begin{aligned}
&\liminf_{m\to\infty} w(b_-^m) - \rho m \geq 0 && (48a)\\
&\Rightarrow \liminf_{m\to\infty} w(b^m) - \rho m \geq 0 && (48b)\\
&\Rightarrow \liminf_{m\to\infty} w(b_+^m) - \rho m \geq 0. && (48c)
\end{aligned}$$

This is obtained by

$$\begin{aligned}
&\liminf_{m\to\infty} w(b_-^m) - \rho m\\
&= w(b_k^n 0) - \rho(n+1) + \liminf_{m\to\infty} w(a^m) - \rho m\\
&= w(b_k^n) - \rho n - \rho + \liminf_{m\to\infty} w(a^m) - \rho m\\
&\leq w(b_k^n) - \rho n + \liminf_{m\to\infty} w(a^m) - \rho m && (49a)\\
&\leq w(b_k^n) - \rho n + (1-\rho) + \liminf_{m\to\infty} w(a^m) - \rho m\\
&= w(b_k^n 1) - \rho(n+1) + \liminf_{m\to\infty} w(a^m) - \rho m && (49b)
\end{aligned}$$

where (49a) equals (48b) and where (49b) equals (48c). The inequalities yield the desired result. ■

[1] Interestingly, in Eggleston's paper, the dimension was not connected to entropy; it was submitted earlier in the same year as Shannon's Mathematical Theory of Communication was published.

References

[1] S. Kahraman, E. Viterbo, and M. E. Çelebi, "Folded tree maximum-likelihood decoder for Kronecker product-based codes," in *Proc. Allerton Conf.*, Oct. 2013, pp. 629–636.

[2] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[3] ——, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Commun. Lett.*, vol. 12, no. 6, pp. 447–449, Jul. 2008.

[4] J. M. Renes, D. Sutter, and S. H. Hassani, "Alignment of polarized sets," in *Proc. IEEE Int. Sym. on Information Theory Proceedings (ISIT)*, Jul. 2015, pp. 2446–2450, extended version available: `arXiv:1411.7925 [cs.IT]`.

[5] E. Abbe, "personal communication," Nov. 2011.

[6] M. Bardet, V. Dragoi, A. Otmani, and J.-P. Tillich, "Algebraic properties of polar codes from a new polynomial formalism," Jan. 2016, `arXiv:1601.06215 [cs.IT]`.

[7] M. Taylor, *Measure Theory and Integration*, ser. Graduate studies in mathematics. American Mathematical Soc., 2006.

[8] S. Haghighatshoar and E. Abbe, "Polarization of the Rényi information dimension for single and multi terminal analog compression," in *Proc. IEEE Int. Sym. on Information Theory Proceedings (ISIT)*, Jul. 2013, pp. 779–783.

[9] E. Abbe and Y. Wigderson, "High-girth matrices and polarization," in *Proc. IEEE Int. Sym on Information Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 2461–2465, `arXiv:1501.06528 [cs.IT]`.

[10] R. Nasser, "Ergodic theory meets polarization. I: An ergodic theory for binary operations," Feb. 2015, `arXiv:1406.2943v4 [cs.IT]`.

[11] ——, "Ergodic theory meets polarization. II: A foundation of polarization theory," Feb. 2015, `arXiv:1406.2949v4 [cs.IT]`.

[12] E. Şaşoğlu, "Polarization and polar codes," *Foundations and Trends® in Communications and Information Theory*, vol. 8, no. 4, pp. 259–381, 2011.

[13] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, 2009.

[14] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.

[15] S. H. Hassani, S. Korada, and R. Urbanke, "The compound capacity of polar codes," in *Proc. Allerton Conf. on Communication, Control, and Computing*, Sep. 2009, pp. 16–21.

[16] M. El-Khamy, H. Mahdavifar, G. Feygin, J. Lee, and I. Kang, "Relaxed channel polarization for reduced complexity polar coding," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, New Orleans, LA, Mar. 2015, pp. 207–212, `arXiv:1501.06091 [cs.IT]`.

[17] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 3rd ed. Chichester: John Wiley & Sons, 2014.

[18] I. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*. New York, NY: John Wiley & Sons, 1974.

[19] S. Albeverio, M. Pratsuivytyi, and G. Torbin, "Topological and fractal properties of real numbers which are not normal," *Bull. Sci. math.*, vol. 129, pp. 615–630, 2005.

[20] T. Kasami, S. Lin, and W. Peterson, "New generalizations of the Reed-Muller codes–I: Primitive codes," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 189–199, Mar. 1968.

[21] P. Delsarte, J. Goethals, and F. M. Williams, "On generalized Reed-Muller codes and their relatives," *Information and Control*, vol. 16, no. 5, pp. 403 – 442, 1970.

[22] S. H. Hassani, K. Alishahi, and R. L. Urbanke, "Finite-length scaling for polar codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5875–5898, Oct. 2014.

[23] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY: McGraw-Hill, 1987.

[24] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. New York, NY: John Wiley & Sons, 1968, vol. 1.

[25] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. New York, NY: Dover Publications, 1972.

[26] H. G. Eggleston, "The fractional dimension of a set defined by decimal properties," *The Quarterly Journal of Mathematics*, vol. os-20, no. 1, pp. 31–36, 1949.